模糊加权的高效鲁棒人体动作视频检索 *

张 涵 [1]，韩 毅 [1,2]，李跃新 [3]

(1. 安阳工学院 计算机科学与信息工程学院，河南 安阳 455000, 2. 华中科技大学 国家数控系统工程技术研究中心，武汉 430000; 3. 湖北大学 计算机工程与通信学院，武汉 430000)

摘　要： 为了提高人体动作视频检索的鲁棒性和效率，提出了一种模糊加权的人体动作视频检索方法。该方法采用 3D Harris 算子检测视频中的时空兴趣点，提取这些兴趣点的梯度信息，构建特征向量；然后采用模糊聚类方法构建聚类特征向量，提高特征向量的抗干扰能力；接着匹配聚类特征向量中的梯度向量对，构建模糊权重矩阵，计算查询视频与数据库中各个视频的相似度；最后在 KTH 数据库上进行视频检索实验，结合精确度、召回率和检索耗时三个指标进行评价，证明该方法的性能最优。

关键词： 视频检索；行为识别；模糊聚类；时空兴趣点；Harris

中图分类号： TP391　　doi: 10.3969/j.issn.1001-3695.2017.08.0956

Efficient and robust video retrieval for human activity with fuzzy weight

Zhang Han[1], Han Yi[1,2], Li Yuexin[3]

(1. *College of Computer Science & Information Engineering*, *Anyang Institute of Technology*, *Anyang Henan* 455000, *China*; 2. *National NC System Engineering Research Center*, *Huazhong University of Science & Technology*, *Wuhan* 430000, *China*; 3. *School of Computer Science & Engineering Hubei University*, *Wuhan* 430000, *China*)

Abstract: In order to improve the robustness and efficiency of video retrieval for human activity, this paper proposed a fuzzy weighted method for human activity video retrieval. This method used 3D Harris operator to detect the spatio-temporal interest points in the video, and extracted the gradient information of these points to construct feature vector to describe video. Then it used fuzzy clustering method to construct uses fuzzy clustering feature vector, to improve the ability of anti-interference of feature vector. And then, it matched pair of gradient vector in the fuzzy clustering feature vectors to construct fuzzy weight matrix, and calculated the similarity between the query video and each video in the database. Finally, it carried out the video retrieval experiment on the KTH database, and carried the evaluation out with three metrics of accuracy, recall and retrieval time, which proved that the performance of this method is the best.

Key Words: video retrieval; behavior recognition; fuzzy clustering; spatio-temporal interest point; 3D Harris

0　引言

随着互联网和多媒体技术的发展，文本、图像、视频等数据呈几何级数增长，这对数据检索技术提出了更多的需求和挑战。文本和图像检索技术已经发展多年，成果很多。而视频检索技术作为一个新的研究方向，近些年受到的关注越来越多。针对视频内容的不同，视频检索技术的研究也有不少差异。本文主要研究人体动作视频的检索技术，此类技术的应用需求更急迫。该视频检索技术是以人体动作行为识别技术为基础的[1]。人体动作行为识别是一种模式识别技术，该技术将不同的人体动作行为描述为不同的模式，通过机器学习等方法构建不同模式的分类器，实现人体动作行为的分类和识别。人体动作行为识别的关键是设计鲁棒的行为描述特征和构建合理的特征分类器[2~4]。尽管人体动作视频检索以人体动作行为识别为基础，但是两者之间还是有明显差异的，主要表现在人体动作视频检索所用的训练样本只有一个，也即查询视频。这样情况下，人体动作视频检索时不能像人体动作行为识别那样通过许多样本的学习来构建分类器，只能通过相似度的计算来推断视频之间的相关性，这样难度相对更大。而且视频检索系统对检索效率的要求一般比较高，因此在设计视频检索算法时还要考虑时效性的问题。目前，人体动作视频检索领域也有一些研究成果，例如文献[5]提出了一种快速的特征对应方法来计算作为相似性度量的匹配花费，然后将相似性度量嵌入到动作检索的多重排序框架中，实现快速的视频检索。文献[6]提出了一种具有 SVM

基金项目： 国家自然科学基金资助项目（64110712）；河南省科技攻关计划资助项目（1721021103）

作者简介： 张涵（1981-），男，河南安阳人，讲师，硕士，主要研究方向为图像处理、信息技术（hanyilaoshi@qq.com）；韩毅，男，（1980-），河南安阳人，讲师，硕士，主要研究方向为计算机应用、智能制造、图像处理；李跃新（1965-），湖北武人，教授，博士，主要研究方向为图像处理、计算机网络等.

相关性反馈的视频检索系统，引入主动学习方法，提高视频检索性能。文献[7]使用视频上下文信息进行建模，引入半监督学习范式启发算法，可以通过很少的样本有效地建模动作模型，提高视频检索准确性。然而现有方法的鲁棒性和时效性都不是很高。为了提高人体动作视频的检索效率和鲁棒性，本文提出了一种模糊加权的人体动作视频检索方法。主要贡献在于引入模糊聚类方法构建聚类特征向量，提高特征向量的抗干扰能力；构建模糊权重矩阵修正欧氏距离测度，提高相似度计算的鲁棒性。同时本文方法的运算复杂度低、运算效率高。

1 人体动作行为识别概述

人体动作行为识别是计算机视觉研究的一个热点，其目标是从一个未知的视频或者图像序列中自动分析其中正在进行的人体动作行为，并识别该动作行为的类别。典型的人体动作行为识别框架[8,9]如图 1 所示。

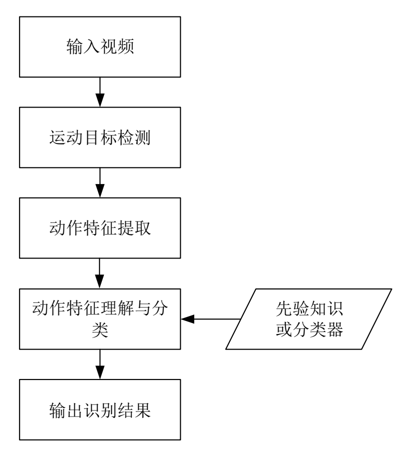


图 1 典型人体动作行为识别流程

首先，考虑到动作行为的运动属性，可以使用运动目标检测技术，从动作视频或者图像中快速提取出感兴趣的目标区域，降低后续动作特征提取与分类的难度。运动目标检测目前已经非常成熟，常用的有帧间差分法、背景减除法和光流法。考虑到人体动作行为的对象是人，因此在运动目标检测的基础上，还可以加入人体的检测方法，例如采用方向梯度直方图特征和支持向量机分类器对运动目标的属性进行检测，辨别运动目标区域内是否存在人体。如果不存在人体，则后续不需要对其进行动作特征的提取和分类操作。在实际的人体动作行为识别过程中，运动目标检测操作并不是必须的，因为有些人体动作行为识别算法是针对整幅图像提取特征的，不便于单独针对运动的人体目标区域提取特征。

动作特征提取是人体动作行为识别的关键。动作特征不仅需要充分体现不同动作之间的差异性，而且还需要适应同一类动作的变化。常用的动作有剪影特征、光流特征、梯度特征、时空特征和深度特征等。剪影特征可以利用目标边界点构建，也可以通过形状上下文构建，如剪影能量图像(silhouette energy image)特征、剪影重构形状(shape-from-silhouette)特征等。剪影特征的优点是受颜色和纹理等不相关特征的影响较小，但受人体遮挡、人体轮廓提取不完整等因素影响较大。光流特征提取的是目标像素点的瞬时运动特征，能够反映目标运动的速度和方向信息，但是受成像条件（如目标距离摄像机远近、摄像机视场等）影响较大，且运算效率偏低。时空特征是针对人体动作在场景中的特殊性构建的描述算子，常用的有时空兴趣点和时空上下文特征，是目前人体动作识别领域应用最多的特征。时空兴趣点是采用一些不关联的点的特征来描述人体的动作信息。常用的时空兴趣点检测方法有 3D-Harris、SIFT 和 Dollar 算子。通过这些算子检测时空兴趣点，然后提取时空兴趣点的特征，如方向梯度直方图特征、时空局部回归核(space-time local regression kernels)特征等。时空上下文特征通过场景上下文、空间上下文和和尺度上下文来提取特征，构建动作场景中动作与人所处环境的相互关系，可采用马可夫逻辑网(Markov logic network)来描述。深度特征是在时空特征的基础上又加入了空间深度信息，这要求采集设备能够获取场景的深度信息，应用比较受限。

动作特征的理解与分类是将提取到的人体动作行为特征与已经学习好的先验知识进行对比，通过模式识别的方式实现人体动作的理解和分类。动作特征理解常用的模型有人体模型和统计模型两类。人体模型包括二维的棍状模型、带状模型以及三维的锥面体模型等，通过分析人体模型的变化来理解人体的动作行为。统计模型包括时空模板、动态规划、状态转移模型等，通过统计视频中特征的变化情况来理解人体的动作行为。动作分类常用到机器学习方法，包括支持向量机、随机森林、神经网络等，通过机器学习方法构建动作特征分类器，实现人体动作行为的分类。

2 人体动作视频检索

给定一个待检索的人体动作视频，人体动作视频检索算法需要依据该视频所提取出的人体动作特征，从视频数据库中寻找具有类似人体动作特征的相关视频，作为检索结果进行输出。人体动作视频检索算法主要包括两个核心环节，一是将视频描述为一个特征向量，要求该特征向量可以描述一类人体动作，并能有效区分不同类别的人体动作；二是计算视频所对应的特征向量之间的相似度，依据相似度检索相关的视频。与人体动作行为识别算法相比，人体动作视频检索算法不需要对行为的类别进行具体的分类，而只需要通过相似度计算寻找与待检索视频相似的动作视频。人体动作视频检索算法更关注检索的时效性和尺度的鲁棒性，这两个方面也是视频检索系统面临的主要挑战。因此，人体动作视频检索算法在借鉴现有人体动作行为识别算法的基础上，还要针对视频检索所面临的这两方面挑战进行适当的优化和改进。本文提出一种的人体动作视频检索方法，基本思路是：采用时空兴趣点的梯度特征构建视频的特征向量；引入模糊聚类方法构建聚类特征向量，提高特征向量的抗干扰能力；构建模糊权重矩阵修正欧氏距离测度，提高相似度计算的鲁棒性。本文方法的实现流程如图 2 所示。其中，

时空兴趣点检测和特征提取部分采用的是人体动作行为识别常用的方法，而相似度计算部分是本文的重要创新。详细描述如下。

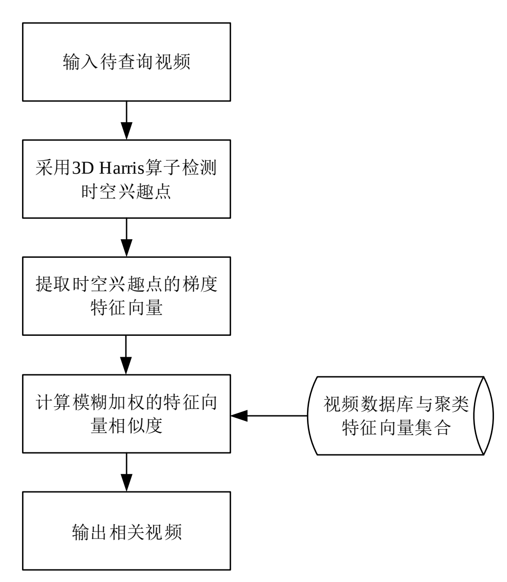


图 2　本文方法实现流程

2.1　时空兴趣点检测

对于每一个视频，首先需要检测时空兴趣点。本文采用常用的 3D Harris 算子来检测时空兴趣点。3D Harris 算子是在空域的 2D Harris 算子的基础上，增加一个时间维度得到的，其检测思想与实现步骤与空域的 2D Harris 算子是相同的。首先，需要进行尺度变换，用尺度空间来表示视频，具体为

$$L(\cdot;\sigma_1;\tau_1)=G(\cdot;\sigma_1;\tau_1)*I \tag{1}$$

其中：σ_1 是空间域的尺度参数；τ_1 是时间域的尺度参数；“*”表示卷积操作；I 是输入视频；G 是高斯核函数，可以表示为

$$G(x,y,t;\sigma_1;\tau_1)=\frac{1}{\sqrt{(2\pi)^3}\sigma_1^{\,4}\tau_1^{\,2}}\exp\left(\frac{-\left(x^2+y^2\right)}{2\tau_1^{\,2}\left(2\sigma_1^{\,2}-t^2\right)}\right) \tag{2}$$

定义

$$E(u,v,p)=\sum_{x,y,t}G(x,y,t)\left[I(x+u,y+v,t+p)-I(x,y,p)\right]^2 \tag{3}$$

按照泰勒公式展开，得到矩阵 M 为

$$M=G(x,y,t)\times\begin{bmatrix}L_x^2 & L_xL_y & L_xL_t\\ L_xL_y & L_y^2 & L_yL_t\\ L_xL_t & L_yL_t & L_t^2\end{bmatrix} \tag{4}$$

进一步地，得到 3D Harris 角点检测算子：

$$\begin{aligned}R&=\det(M)-k\times(\mathrm{trace}(M))^3\\&=\lambda_1\lambda_2\lambda_3-k\times(\lambda_1+\lambda_2+\lambda_3)^3\end{aligned} \tag{5}$$

其中：det(M)表示矩阵 M 的行列式；trace(M)表示矩阵 M 的迹；k 为常数，常取 0.04~0.06；λ_1、λ_2 和 λ_3 是矩阵 M 的特征值，具体的实现过程详见文献[10]。

2.2　特征向量提取

本文采用时空兴趣点的梯度信息来构建描述视频的特征向量。当检测到一个时空兴趣点之后，使用一个时空立方体，将梯度描述子作用于每一个立方体，计算时空兴趣点的主要运动方向和尺度。然后将时空立方体每一个轴上的梯度描述子的结果串联在一起，构成一个特征向量，称为时空兴趣点的梯度向量。具体的实现过程详见文献[11]。需要指出的是，本文不需要对梯度向量进行降维操作，这样可以避免造成特征的损耗。而且本文在相似度计算环节会通过聚类操作来降低特征向量比对的计算复杂度，与文献[11]的降维操作一样可以实现提高运算效率的目标。这样，得到视频中所有时空兴趣点的梯度向量之后，即可构建视频的特征向量描述子。假设 v_i 表示任一视频，该视频提取的到时空兴趣点的数量为 n_i，对应的特征向量可以表示为 $F_{v_i}=\left\{f_{v_i,j}\,\middle|\,j=1,2,\cdots,n_i\right\}$。其中：$f_j$ 表示第 j 个时空兴趣点提取的特征向量。

2.3　相似度计算

对于每一个视频提取的特征向量，本文首先采用模糊 k 均值聚类算法，对特征向量中各个时空兴趣点的梯度向量进行聚类，提高特征向量的抗干扰能力，实现过程可以参考文献[12]。假定聚类中心数量为 c，那么特征向量 F_{v_i} 聚类之后可以表示为 $F_{v_i}^{(c)}=\left\{f_{v_i,j}^{(c)}\,\middle|\,j=1,2,\cdots,c\right\}$。其中：$f_{v_i,j}^{(c)}$ 表示聚类之后视频 v_i 的特征向量中的第 j 类梯度向量。

当完成特征向量的模糊聚类之后，计算查询视频的聚类特征向量与数据库中各个视频的聚类特征向量之间的相似度，并按照相似度的大小进行排序。两个聚类特征向量之间的相似度越大，说明它们对应的两个视频越相关。

通常，两个视频之间的相似度可以通过其对应的两个聚类特征向量之间的距离来反映，距离越近，说明两个视频之间的相似度越大。本文也基于这一思想设计视频相似度的计算公式。具体到聚类之后的特征向量，考虑到该特征向量包含了 c 类梯度向量，而同一类视频在不同的采集片段可能存在尺度、位置、时序等方面的差异，这种情况下，两个聚类特征向量中的 c 类梯度向量一般不是一一对应的。因此，本文先计算两个聚类特征向量中任意两类梯度向量之间的距离，选择拥有最小距离的两类梯度向量构建一个梯度向量对，通过这样的方式将 c 类梯度向量的对应关系确定下来。同时，考虑到尺度、位置等因素的影响可能导致梯度特征向量在模值上存在较大差异，譬如离摄像机越近的目标梯度模值越大，而离摄像机越远的目标梯度模值越小。因此，本文引入模糊权重对梯度向量进行归一化，然后归一化的距离来计算两个聚类向量之间的相似度。下面详述视频相似度的计算方式。

假设 v_i 表示待查询视频，v_j 表示数据库中某一视频，视频 v_i 和 v_j 对应的聚类特征向量分别记为 $F_{v_i}^{(c)}$ 和 $F_{v_j}^{(c)}$。那么，对于

聚类特征向量对 $\left(F_{v_i}^{(c)}, F_{v_j}^{(c)}\right)$，$F_{v_i}^{(c)}$ 的第 p 类梯度向量和 $F_{v_j}^{(c)}$ 的第 q 类梯度向量之间的距离可以记为 $d_{v_i,v_j}(p,q)$。本文采用欧氏距离测度，于是有

$$d_{v_i,v_j}(p,q)=\left\|f_{v_i,p}^{(c)}-f_{v_j,q}^{(c)}\right\|_2 \tag{6}$$

$F_{v_i}^{(c)}$ 的第 p 类梯度向量和 $F_{v_j}^{(c)}$ 的第 q 类梯度向量之间的模糊权重可以记为 $w_{v_i,v_j}(p,q)$,计算公式为

$$w_{v_i,v_j}(p,q)=\left(\sum_{k=1}^{c}\left(\frac{d_{v_i,v_j}\left(f_{v_i,p}^{(c)}, f_{v_j,q}^{(c)}\right)}{d_{v_i,v_j}\left(f_{v_i,p}^{(c)}, f_{v_j,k}^{(c)}\right)}\right)^{\frac{2}{f-1}}\right)^{-1} \tag{7}$$

其中：f 为模糊常量。在本文中，f 取值为 0.3。

这样，聚类特征向量对 $\left(F_{v_i}^{(c)}, F_{v_j}^{(c)}\right)$ 中的 c 类梯度向量的模糊权重可以构建一个 $c\times c$ 的模糊权重矩阵，记为

$$W_{v_i,v_j}=\begin{bmatrix} w_{v_i,v_j}(1,1) & w_{v_i,v_j}(1,2) & \cdots & w_{v_i,v_j}(1,c) \\ w_{v_i,v_j}(2,1) & w_{v_i,v_j}(2,2) & \cdots & w_{v_i,v_j}(2,c) \\ \vdots & \vdots & \ddots & \vdots \\ w_{v_i,v_j}(c,1) & w_{v_i,v_j}(c,2) & \cdots & w_{v_i,v_j}(c,c) \end{bmatrix} \tag{8}$$

聚类特征向量 $F_{v_i}^{(c)}$ 中的第 p 类梯度向量所对应的聚类特征向量 $F_{v_j}^{(c)}$ 中的梯度向量序号可以表示为

$$p^*=\min_{q=C}\left(w_{v_i,v_j}(p,q)d_{v_i,v_j}(p,q)\right) \tag{9}$$

其中：C 表示序号 q 的取值集合。这里需要说明的是，在从聚类特征向量 $F_{v_j}^{(c)}$ 的 c 类梯度向量中寻找与聚类特征向量 $F_{v_i}^{(c)}$ 的第 p 类梯度向量相匹配的梯度向量时，如果某一类梯度向量已经被选择为对方的梯度向量对，那么下一次寻找过程中该梯度向量就不再参与运算。也即第一次寻找时，$C=\{i|i=1,2,\cdots,c\}$；第二次寻找时，$C=\{i|i=1,2,\cdots,c;i\neq 1^*\}$。这里，$1^*$ 表示聚类特征向量 $F_{v_i}^{(c)}$ 的第 1 类梯度向量从聚类特征向量 $F_{v_j}^{(c)}$ 中找到的匹配梯度向量的序号。依此类推，每次搜索时集合 C 的元素数量都会减少 1。

这样，聚类特征向量 $F_{v_i}^{(c)}$ 的第 p 类梯度向量和聚类特征向量 $F_{v_j}^{(c)}$ 的第 p^* 类梯度向量组成一个 $F_{v_i}^{(c)}$ 和 $F_{v_j}^{(c)}$ 之间的一个梯度向量对 $\left(f_{v_i,p}^{(c)}, f_{v_j,p^*}^{(c)}\right)$。类似地，可以得到 $F_{v_i}^{(c)}$ 和 $F_{v_j}^{(c)}$ 之间的其余 c-1 个梯度向量对。

于是，聚类特征向量 $F_{v_i}^{(c)}$ 和与聚类特征向量 $F_{v_j}^{(c)}$ 之间的距离可以用 c 个梯度向量对之间模糊距离的平均值来表示为

$$d_{v_i,v_j}=\frac{1}{c}\sum_{k=1}^{c}\left(w_{v_i,v_j}(k,k^*)d_{v_i,v_j}(k,k^*)\right) \tag{10}$$

距离越近，明两个视频之间的相似度越大。因此，视频 v_i 和 v_j 之间的相似度可以表示为

$$s_{v_i,v_j}=\frac{1}{d_{v_i,v_j}+\varepsilon} \tag{11}$$

其中：ε 为一个非常小的正数，用于避免除数为零。在本文中取值为 0.000 01。

2.4　相关视频输出

对于待检索的视频，按照上述步骤计算其与数据库中所有视频之间的相似度，然后按照相似度从大到小的顺序进行排列，相似度越大，对应的视频之间的相关性越强。在视频检索时，往往需要检索的视频数量不止一个。通常需要设置一个查询余量参数，记为 U。也就是说，每一次检索都输出 U 个相关视频作为检索结果。在本文中，相似度降序排列之后，输出前 U 个相似度所对应的数据库中的 U 个视频作为检索结果。

3　实验分析

由于人体动作视频检索领域目前没有专门公开的测试数据库，所以本章使用人体动作行为识别领域常用的 KTH 数据库进行视频检索实验。该数据库是人体动作行为识别领域的一个大规模数据库，包含 2 391 个视频片段，每一个视频片段中包含一个人体重复的单一动作，共有 6 种动作，分别是行走、慢跑、跑、拳击、挥手和拍手，如图 3 所示。动作的执行者共有 25 人，视频拍摄场景有 4 种。视频中人体的着装、尺度都有变化。视频的分辨率为 160×120，视频的帧率为 25 fps。



图 3　KTH 数据库动作示例

本文在 KTH 数据库上进行人体动作视频的检索实验，统计视频检索性能指标。这里参考图像检索领域常用的精确度（precision）和召回率（recall）两个指标来评价视频检索算法的准确性，同时采用检索耗时指标来评价视频检索算法的运算效率。在本文实验中，每一类行为都选择数据库中的前 100 个

视频作为查询视频，其他视频作为检索数据库。查询余量参数 U 统一设置为 20。下面首先详细测试本文方法的检索准确性，然后对比不同的视频检索方法综合评价本文方法性能。

3.1 本文方法的检索准确性测试

在上文的介绍中可以发现，本文方法还有一个参数没有赋值，即聚类中心数量 c。该参数对视频检索的准确性是有较大影响的，通常情况下，聚类中心数量越大，特征向量划分越细，不同行为之间的区分能力越强，但相同行为之间的容错能力越差。下面通过对比不同参数 c 条件下的平均精确度和平均召回率指标来选择最优的聚类中心数量参数。图 4 给出了测试结果。由图 4 可以看出，当参数 c 取值为 6 时，平均精确度和平均召回率都是最大的。因此，本文方法中参数 c 取值为 6。此时，本文方法的平均精确度和平均召回率分别是 78.1%和 73.8%。

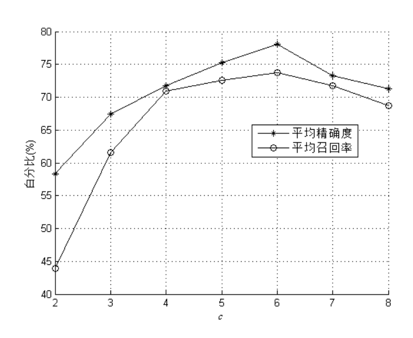


图 4 参数 c 取值实验结果

3.2 不同方法的视频检索性能对比

人体动作视频检索是近些年才流行的研究课题，目前这方面的研究成果不多。文献[5~7]是目前人体动作视频检索领域比较典型的检索方法。本文选择这三种方法进行对比实验，实验数据库、检索视频、查询余量参数以及实验计算机平台都与本文方法相同，计算机平台性能为 Intel I5 CPU、16 GB RAM、Windows 7 64 bit、Visual Studio 2013。实验结果见表 1。其中，平均值是指平均精确度和平均召回率的平均值，用于反映两者的综合指标。平均检索耗时的单位是 s。

表 1 不同方法的视频检索指标

方法	平均精确度	平均召回率	平均值	平均检索耗时
文献[5]方法	67.3%	59.4%	63.4%	73
文献[6]方法	71.6%	72.9%	72.3%	166
文献[7]方法	**78.8%**	69.7%	74.3%	231
本文方法	78.1%	**73.8%**	**76.0%**	**59**

由表 1 可以看出，在同等条件下，本文方法的平均召回率指标是四种方法中最高的，尽管本文方法的平均精确度指标略低于文献[7]方法，但是从两者的均值来对比，本文方法仍是四种方法中最高的。也就是说，综合平均精确度和平均召回率两个指标，本文方法优于其他三种方法。而且本文方法的平均检测耗时明显低于其他三种方法，这说明本文方法的检索效率明显高于其他三种方法，这对于视频检索系统而言是非常有意义的。

4 结束语

人体动作视频检索是一项较新的研究课题，在互联网、多媒体领域有着广阔的应用前景。鉴于视频检索系统对检索效率的较高要求，本文提出了一种高效的人体动作视频检索方法。该方法在传统的基于时空兴趣点和梯度信息提取特征向量的基础上，引入模糊聚类方法构建聚类特征向量，提高了特征向量的抗干扰能力。同时，在检索过程中先匹配聚类特征向量中的梯度向量对，再构建模糊权重矩阵，对欧氏距离测度进行修正，提高相似度计算的鲁棒性。通过实验对比，验证了本文方法不仅精确度和召回率综合指标高，而且检索耗时少，检索效率高。

参考文献：

[1] 梁俊杰, 熊亚军, 余敦辉. 一种基于本体的视频检索技术研究 [J]. 计算机工程与科学, 2015, 37 (10): 1940-1946.

[2] Chaaraoui A A, Climent-Pérez P, Flórez-Revuelta F. Silhouette-based human action recognition using sequences of key poses [J]. Pattern Recognition Letters, 2013, 34 (15): 1799-1807.

[3] Chen C, Jafari R, Kehtarnavaz N. Improving human action recognition using fusion of depth camera and inertial sensors [J]. IEEE Trans on Human-Machine Systems, 2015, 45 (1): 51-61.

[4] Chaaraoui A A. Evolutionary joint selection to improve human action recognition with RGB-D devices [J]. Expert Systems with Applications An International Journal, 2014, 41 (3): 786-794.

[5] Tang J, Shao L, Zhen X. Human action retrieval via efficient feature matching [C]// Proc of IEEE International Conference on Advanced Video and Signal Based Surveillance. 2013: 306-311.

[6] Jones S, Shao L, Du K. Active learning for human action retrieval using query pool selection [J]. Neurocomputing, 2014, 124 (2): 89-96.

[7] Jiang Y G, Li Z, Chang S F. Modeling scene and object contexts for human action retrieval with few examples [J]. IEEE Trans on Circuits & Systems for Video Technology, 2011, 21 (5): 674-681.

[8] 李瑞峰, 王亮亮, 王珂. 人体动作行为识别研究综述 [J]. 模式识别与人工智能, 2014, 27 (1): 35-48.

[9] 冯家更, 肖俊. 视点无关的行为识别综述 [J]. 中国图象图形学报, 2013, 18 (2): 157-168.

[10] Sipiran I, Bustos B. Harris 3D: a robust extension of the Harris operator for interest point detection on 3D meshes [J]. The Visual Computer, 2011, 27 (11): 963-976.

[11] Fang X, Tian Y, Wang Y, et al. Pair-wise event detection using cubic features and sequence discriminant learning [C]// Proc of IEEE International Conference on Multimedia and Expo. 2013: 1-6.

[12] Cebeci Z, Yildiz F. Comparison of K-means and fuzzy C-means algorithms on different cluster structures. [J]. Journal of Agricultural Informatics, 2015, 6 (3): 13-23.